PI NAME: Eleftherios Papoutsakis

Department of Chemical Engineering and the Delaware Biotechnology Institute,

University of Delaware

ONR award number: N000141010161

ONR Award Title: Engineering Complex Microbial Phenotypes with continuous genetic integration and plasmid based multi-gene library

a. Scientific and Technical Objectives

To develop tools & strategies to facilitate the development of complex phenotypes in microbial cells by combining DNA from different organisms in order to develop desirable complex traits, such as tolerance for chemicals, such as ethanol (EtOH) and 1,2,4-butanetriol (BT), which are used as "testbeds" for the proof of the proposed concept. BT is a chemical of interest to Navy for possible production by a biological process using metabolically engineered cells.

Objective 1: Build complex phenotypes by accelerated evolution to generate "new" genomes in order to enable selection of desired phenotypes. The goal is to enlarge the *Escherichia coli* genome by selective integration of genes from strains that have desirable traits. Here, we aim to enlarge the *E. coli* genome using *Lactobacillus plantarum* genes to build cells tolerant to EtOH and BT. *L. plantarum* is an organism with established high tolerance to alcohols and solvents more broadly.

Objective 2: Build a stress-response (or chaperone) system that can be customized for tolerance to chosen toxic chemicals, such as EtOH and BT.

b. Approach

1. Integrate heterologous (here: *L. plantarum*; abbreviated as *L. pl*) DNA into the *E. coli* chromosome while selecting for insertions that enhance ethanol tolerance (which is used as a generic desirable trait to provide a proof of the concept). An antibiotic gene marker was inserted into the *E. coli* chromosome to be used as a target for homologous integration of a second antibiotic gene marker along with ~ 3 kb *L. pl* DNA fragments. The second antibiotic gene marker was the target for homologous integration of another antibiotic gene marker along with additional ~3 kb *L. pl* DNA fragments. This approach can be repeated for integration of additional DNA fragments into the *E. coli* chromosome. The screening for tolerance is used so that the integrated heterologous DNA contributes to the development of the desirable phenotype.

2. Build library of HSP proteins to screen for the impact of co-expressed HSPs on the development of tolerance. We will utilize the MultiSite Gateway® (Invitrogen) to clone sets of three HSPs into a single expression plasmid. For proof of concept, a core heat shock protein

Report Documentation Page

Form Approved
OMB No. 0704-0188

1. REPORT DATE	2. REPORT TYPE	3. DATES COVERED
2010	**Annual**	**-**

4. TITLE AND SUBTITLE	5a. CONTRACT NUMBER
Engineering Complex Microbial Phenotypes with continuous genetic integration and plasmid based multi-gene library	5b. GRANT NUMBER
	5c. PROGRAM ELEMENT NUMBER
6. AUTHOR(S)	5d. PROJECT NUMBER
	5e. TASK NUMBER
	5f. WORK UNIT NUMBER
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) **University of Delaware**	8. PERFORMING ORGANIZATION REPORT NUMBER
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)	10. SPONSOR/MONITOR'S ACRONYM(S)
	11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release, distribution unlimited

13. SUPPLEMENTARY NOTES
The original document contains color images.

14. ABSTRACT

15. SUBJECT TERMS

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			
unclassified	**unclassified**	**unclassified**	**SAR**	**13**	

(HSP) system (the autologous GroESL system) was overexpressed in *E. coli* aiming to provide tolerance to ethanol and 1,2,4-butanetriol.

c. Concise Accomplishments

1. Proof-of-concept that we can integrate heterologous DNA into *E. coli* in an iterative process that can be used to continuously enlarge the *E. coli* genome by integrating *L. plantarum* (*L. pl* or any other) DNA. 1) *E. coli* K12 and *E. coli* Epi300 T1R Tn5 disruption libraries were obtained by integrating heterologous DNA together with Km^R (kanamycin resistance) gene randomly inserted in the *E. coli* chromosome. The generated K12 library had 2000 clones and the Epi300 T1R library 20000 clones. 2) We have successfully inserted a Cm^R (chloramphenicol resistance) gene along with 3 kb *L. pl* DNA fragments into the Km^R gene in the Epi300 Tn5 library (1000 clones): this is the Epi300 Clp library. 3) 2nd step clones were selected to resist up to 5.5% ethanol. Selected clones were used for 3rd step DNA integration. 4) 3rd step recombinants were obtained with Sp^R (spectinomycin resistance) gene along with a second 3 kb *L. pl* DNA fragment integrated into the Cm^R gene of the Epi300 Clp library (1000 clones).

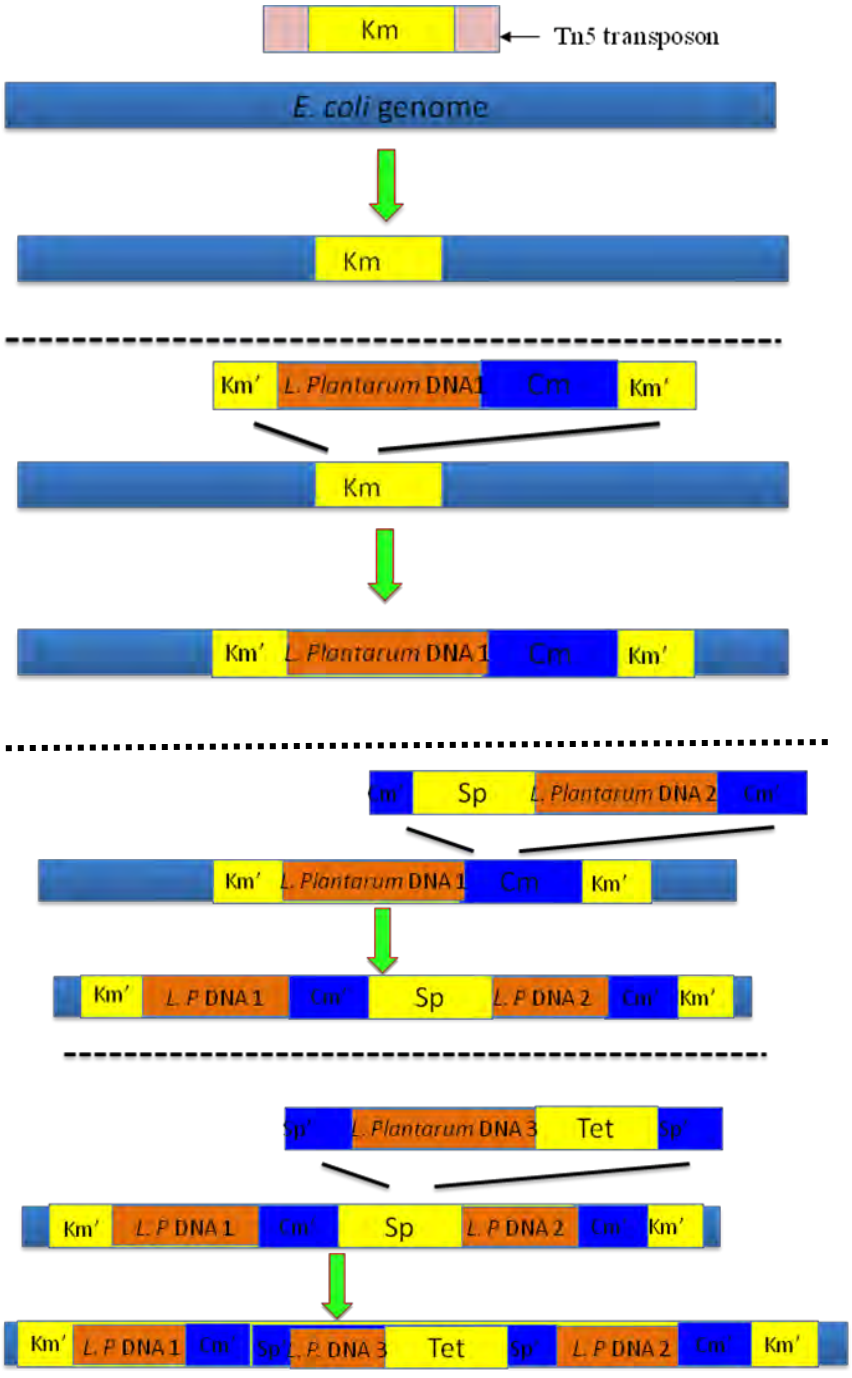
2. Proof-of-concept that overexpression of HSPs generates a strong tolerant phenotype against ethanol and 1,2,4-butanetriol. Progress made in developing a system to screen for beneficial effects of overexpressing multiple HSPs.

d. Expanded Accomplishments

Objective 1: Building complex phenotypes by accelerated evolution to generate "new" genomes

This aim is to develop an approach based on DNA genomic-libraries for generating random gene knockouts and DNA insertions into the *E. coli* chromosome as a means for building complex phenotypes. The idea is to use homologous recombination for random integration into and disruption of the *E. coli* genome by a cassette consisting of an antibiotic resistance marker along with a heterologous (here: *L. plantarum*) genomic DNA fragment. We first created a DNA library using genomic DNA from an organism whose DNA we desire to "sample" for integration into *E. coli*. After a round of random DNA integration (by a mechanism we discuss below) into the *E. coli* chromosome, we created a diverse *E. coli* population which has integrated one DNA segment from the heterologous DNA library. We subsequently integrated additional heterologous DNA in a process that can repeated many times in cycles. We have done so by targeting as an integration site the antibiotic selection marker used in the second round of integration and by using a different selection marker. In a 3rd round of chromosomal integrations, we targeted the integration in that 2nd antibiotic marker (the 1st antibiotic marker was already largely deleted from the chromosome). This process can be repeated until the desired number of chromosomal inserts is achieved in order to enlarge the *E. coli* genome and increase the diversity of the inserted DNA. After each cycle of DNA integration or following three or more rounds of DNA integration, we can screen for a desirable phenotype to select for DNA insertions that promote the development of the desirable phenotype, which here was chosen to be tolerance to various simple and complex alcohols. We hypothesize that the combination of random gene

disruptions coupled with multiple rounds of insertion of heterologous DNA will generate a large population of "different genomes" which can screened for complex phenotypes. The continuous integration steps are illustrated in Figure 1 and explained below.



1st Round: Random disruption of the chromosome with Tn5 transposon containing Km gene.

2nd Round: Disrupting Km to insert the *1st L. plantarum* DNA fragment with a Cm marker in the chromosome.

3rd Round: Disrupting the Cm gene to insert heterologous DNA with Sp marker in the chromosome. Repeat this step for continuous integration.

4th Round: Disrupting the Sp gene to insert heterologous DNA with Tet marker in the chromosome. Repeat this step for continuous integration.

Figure 1. steps of the 4 rounds of continuous chromosome integration

1st step. Initial chromosomal disruption and insertion of genomic DNA: Round 1 (Fig. 1). The Tn5 transposome (Epicentre) was used to generate the *E. coli* K12 and an *E. coli* Epi300 knock-out (KO) library with the Km^R gene randomly inserted on the chromosome. The K12 library contains about 2000 individual KO strains and the Epi300 library contains about 20000 KO strains. Out of the K12 library, 6 colonies were randomly selected to identify the chromosomal insertion sites to demonstrate the diversity of the library. As shown in Table 3, all 6 strains had the Tn5 cassette inserted at different locations on the chromosome.

Table 1. The insertion locations of the Tn5 cassettes on K12 chromosome.

K12 Tn5 mutants	Tn5 cassette insertion location
1	3746904
2	496291
3	4149749
4	3860285
5	3119251
6	4060121

2nd step. Utilizing the first antibiotic resistance marker as a gateway for the insertion of more genomic DNA into the *E. coli* chromosome: Round 2, 3, etc of DNA chromosomal integrations (Fig. 1). After the 1st Round of disruption /integration into the chromosome, the selection marker serves as the target region of homology so that additional DNA can be inserted into the antibiotic resistance gene. In this case, a linear DNA fragment as illustrated in Figure 1, round 2 was constructed with a Cm^R gene along with a random 3 kb *L. pl* DNA, which was flanked by 70 kb 5' and 3' Km^R gene segments for homologous recombination. To construct this linear DNA in large amounts and to prevent plasmid contamination for the transformation, a *rep*101-*ori*101 based plasmid was chosen as a plasmid backbone. These plasmids are not stable in *E. coli* at temperatures higher than 30 C. As described in Figure 4, long PCR primers with the Km^R gene sequences and cloning sites were designed to clone the Cm^R gene. The PCR product was ligated to *rep*101-*ori*101 based plasmid backbone. The plasmid was digested by *Bgl* II, and ligated to 3 kb *L. pl* genomic DNA. The constructed plasmid library was then cut by *Fse* I to obtain the linear DNA for electroporation into the *E. coli* mutant library constructed in the 1st round (Figure 1). The same protocol was used to construct linear DNA containing a Sp^R gene or Tc^R (tetracycline reistance) gene to integrate into the antibiotic resistance gene present in the parent strain library for the 3rd and 4th rounds of DNA integration (Figure 1). Multiple rounds of insertions can be thus performed.

The 2nd round integration library with the CmR gene and the first 3 kb *L.pl* genomic DNA inserted into the chromosome of *E. coli* Epi300 host were successfully established with about 1000 individual strains. The library was cultured at 42 C for 27 generations to eliminate plasmid contamination. Although the individual strains of this library could not reach 1x coverage of the total *L.plantarum* genome, this small-scale library established here was for proof-of-concept only. The coverage of the integration library will be expanded by optimizing the efficiency of the experimental process or by combining more transformations together.

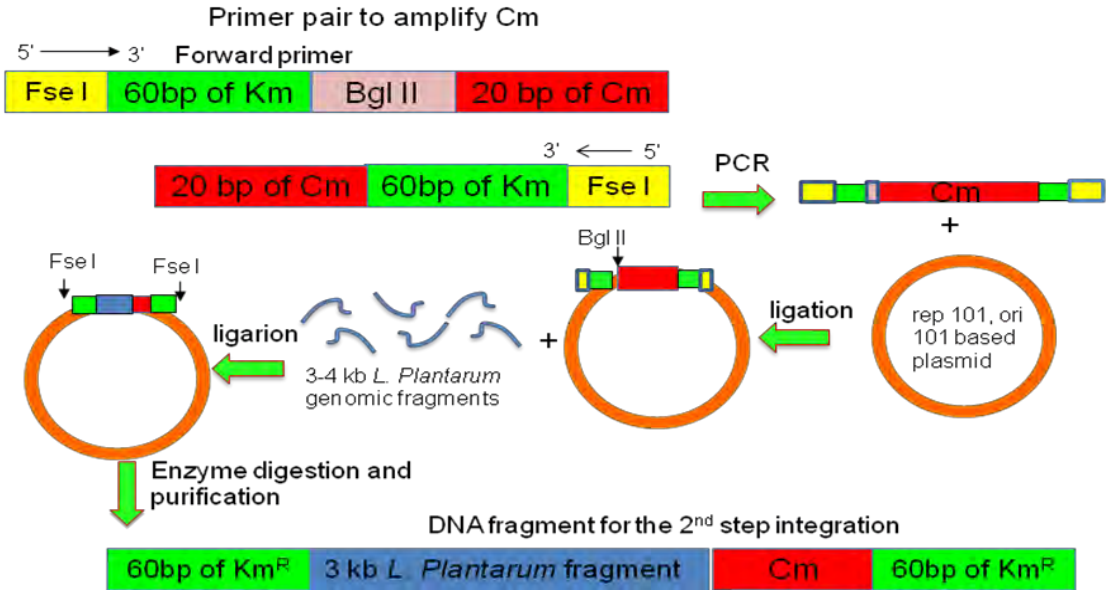


Figure 2. Steps to construct the linear DNA for 2nd step integration of Cm^R gene along with the 1st *L. plantarum* DNA fragment.

Table 2. The sequences of *L. pl* WCFS1 DNA fragments integrated in the original (i.e., the library without selection for ethanol tolerance) 2nd round integration library.

Sample number	Start position	End position	length	Genes
1	1899576	1894976	4600	Conserved hypothetical protein
2	592144	596972	4828	prophage lp1
3	1727178	1721789	5389	ABC transporter, clp protease
4	1931776	1927192	4584	cell division protein ftsw, pyruvate carboxylase
5	1727180	1721787	5393	ABC transporter, clp protease

Table 3. The sequences of *L. pl* WCFS1 DNA fragments integrated in the 5.5% ethanol enriched 2nd round integration library.

Sample number	Start position	End position	length	Gene(s)
1	**1126197**	**1123026**	**3171**	**hypothetical protein, priming glycosyltransferase**
2	**1727181**	**1721803**	**5378**	**ABC transporter clp protease**
3	**592142**	**596970**	**4828**	**prophage lp1**
4	**592148**	**596970**	**4822**	**prophage lp1**
5	**592136**	**596970**	**4834**	**prophage lp1**
6	**3001950**	**2996511**	**5439**	**Prophage lp4 protein 8, purine-cytosine transporter**

To confirm the recombinant event illustrated in Round 2 of Figure 1, whereby the Cm^R gene along with a 3kb L. *pl* DNA fragment are inserted into the Km^R gene, colony PCR was employed. The results are illustrated in Figure 2. The amplified integrated DNA fragments were sequenced to be 4 kb *L. pl* DNA. Both the original (i.e., the library without selection for ethanol tolerance) *E. coli* library and the 5.5% ethanol enriched library were analyzed. The sequencing results of inserted *L. pl* DNA fragments are summarized in Table 2 and 3.

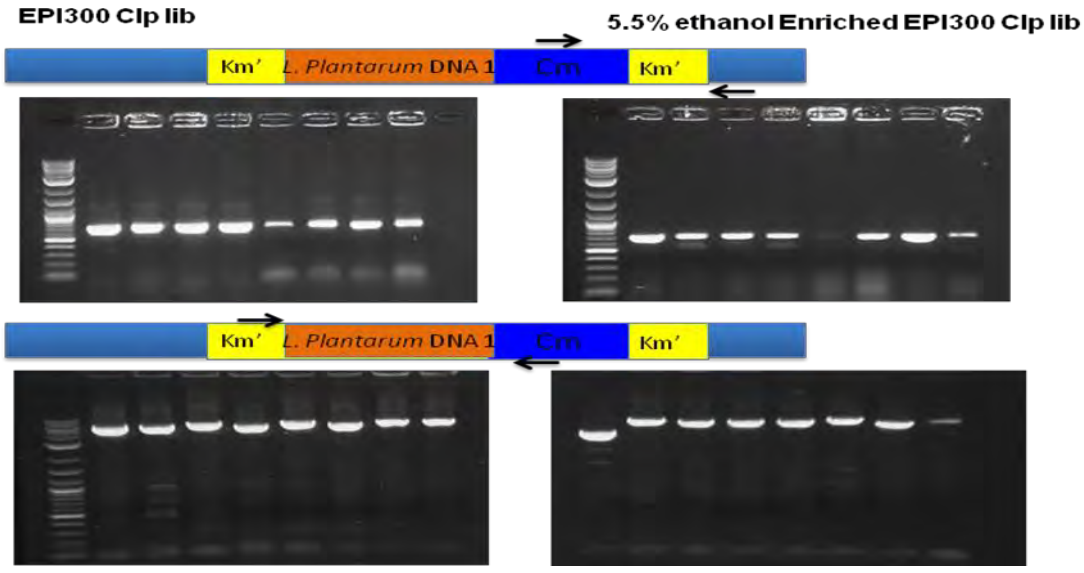


Figure 2. Colony PCR to confirm that the Cm^R gene along with a 3 kb *L. pl* DNA fragment inserted a Km^R gene, and to identify the integrated 3 kb *L. pl* DNA fragment. The position of the primers used for confirmation are represented by the arrow symbol placed on diagrammatically shown genetic segments.

It was determined by the results that the small scale 2nd step integration library was successfully constructed with expected genetic features. Because there was a 27-generation culturing to eliminate plasmid contamination, some of the inserted *L. pl* DNA was preferentially enriched in the library, such as the 5390 bp fragment located on *L. pl* WCFS1 chromosome from 1721803 to 1727181. This bias will be eliminated by using non-replicating plasmids to amplify the linear DNA for integration.

3rd step. Screening of the generated *E. coli* integrant libraries for a desirable complex phenotype. Multiple rounds of insertion of heterologous DNA will be performed to create a library of *E. coli* integrants. Those cells will be screened by exposure to selective or differential conditions (exposure to a stress vs. not; here: ethanol, or butanol, or 1,2,4-butanetriol). The assumption is that the gene disruptions and DNA inserts will create complex phenotypes that increase tolerance and allow for preferential growth under the applied stress. Thus, the population will eventually be taken over by cells carrying inserts that provide a selective advantage. Briefly, the collection of cells will be stressed until the stationary phase of cultures, and will then be transferred into fresh media with higher stress concentrations. The repeated stress will enrich for tolerant strains, which can be further characterized at a later stage. We designed two parallel selection protocols towards the desired phenotypes as illustrated in Figure 3. In protocol (process) 1, stress for selection is applied after all 3 (or more) heterologous DNA fragments are integrated into the *E. coli* chromosome. In protocol (process 2), a selection stress is applied after each round of genetic integration. The enriched library after each integration is used as the strain library for the next round of genetic integration. For example, the 5.5% ethanol enriched 2nd round integration library created in step 2 is used as the strain library for the 3rd round of DNA integration as illustrated in Figure 1. When the library was cultured with ethanol stress, some of the inserted *L. pl* genetic elements were enriched as expected (Table 3). For example, the 4830 bp *L. pl* genetic fragment containing prophage Lp1 was preferentially enriched in the population of the 2nd round integration library after ethanol stress. These enriched *L. pl* genetic elements might be able to contribute to tolerance the ethanol stress in *E. coli*.

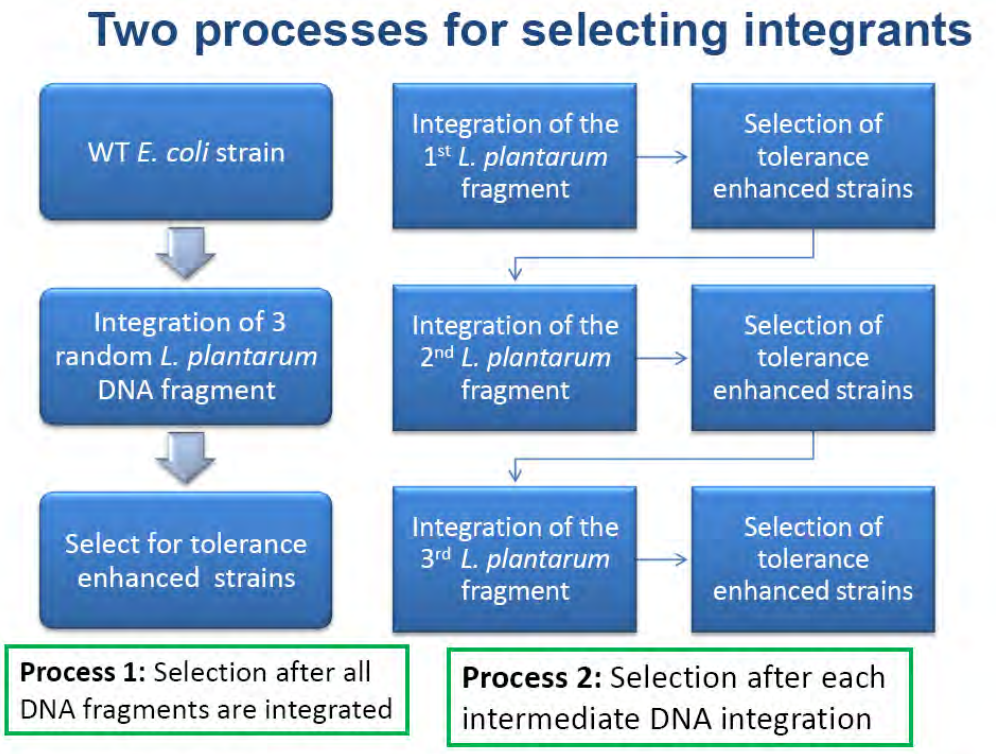


Figure 3. Two parallel selection protocol for developing the desired complex phenotypes.

So far, we have successfully constructed and genetically characterized a small-scale 2nd round integration library of *E. coli*. The 3rd round genetic integration library was also constructed and currently under characterization. An ethanol tolerance selection protocol was applied and we demonstrated that the inserted *L. plantarum* genetic elements apparently contributed to the host tolerant phenotype to ethanol stress. These data constitute a proof-of-concept for the ideas summarized in Figures 1 and 3.

Objective 2: Build a stress-response (or chaperone) system that can be customized for tolerance to chosen toxic chemicals

First, we aimed to provide evidence for the hypothesis that oeverexpression of select HSP proteins may generate a strong tolerant phenotype against ethanol and 1,2,4-butanetriol. To do so we overexpressed the *E. coli* HSP genes *groES* and *groL* and showed that the resulting strain has increased tolerance to the solvents ethanol and 1,2,4-butanetriol. These autologous *E. coli groESL* genes were cloned into the plasmid pACYC184 under their natural promoter and transformed into the 10-β strain of *E. coli*. The strain was then used in tolerance experiments using as control the strain containing the empty pACYC184 vector.

After induction of the stress response genes through growth of the inoculums in 3% ethanol, the rate of growth and growth extents were measured by measuring cell densities (by optical density at 600 nm) (Figures 4 and 5). Both the control and the *groESL* overexpression strain

demonstrated similar growth without solvent stress. However, growth in the presence of the toxic solvents caused significant slower growth in the control strain after induction by 3% ethanol.

The increased tolerance (Figures 4 and 5) observed with the higher expression levels of *groESL* genes suggest that through expression of multiple HSP (stress-response) genes, it is possible to increase the tolerance of a microorganism to solvent stress. This is proof-of-principle for generating increased tolerance phenotypes through over-expression of multiple HSP genes, and suggests that a customizable phenotype can be generated through screening a library of HSP genes.

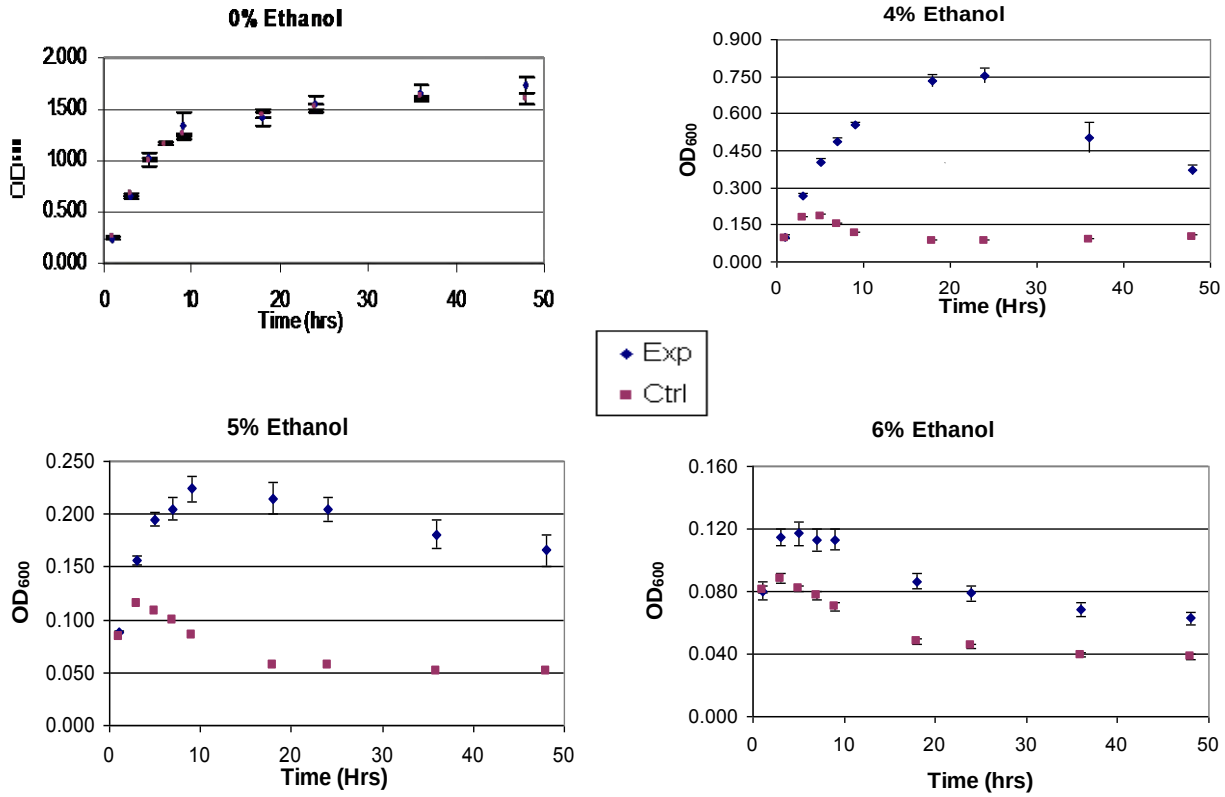


Figure 4. Growth of *groESL* overexpression strain (vs. the plasmid control strain) in the presence of increasing concentrations of ethanol after induction through growth of the inoculum in 3% ethanol.

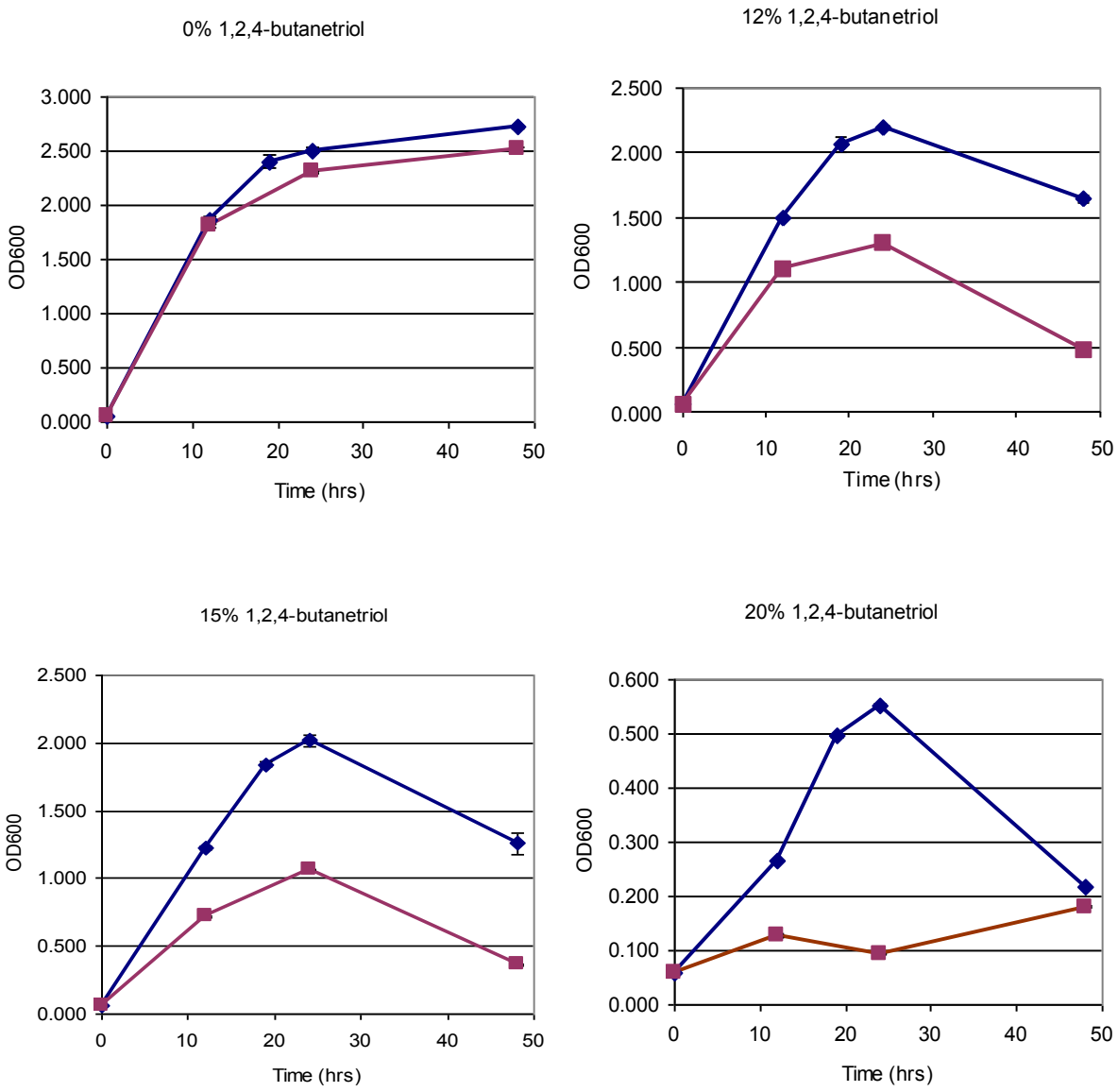


Figure 5. Growth of *groESL* overexpression strain (vs. the plasmid control strain) in the presence of increasing concentrations of 1,2,4-butanetriol after induction through growth of the inoculum in 3% ethanol.

e. Work Plan

Objective 1:
We will follow the protocol we established in order to construct the *E. coli* chromosomal integration library with 3 random genetic fragments from *L. plantarum*. The two selection methods as described above will be used to select for *E. coli* strains with enhanced tolerance to ethanol and 1,2,4-butanetriol. These selected strains will be analyzed in order to identify the

heterologous *L. pl* genetic elements, whose protein products resulted in the tolerant phenotype. Several issues are to be addressed in next year. First, the current method employed temperature sensitive plasmid in order to produce linear DNA for transformation. Substantial sub-culturing was required in order to eliminate the plasmid contamination accompanying the linear-DNA transformation. This likely introduced a bias in our *E. coli* library. A non-replicating plasmid method is currently being developed for solve this problem of library bias. Second, a much larger *E. coli* library of is to be constructed in order to increase the number of possible combinations of the three or more *L. pl* DNA fragments integrated into the *E. coli* chromosome. This is likely to increase the probability of identifying combinations of *L. pl* genomic fragments that provide tolerance to toxic chemicals. The address this issue, the efficiency of transformation and recombination will be increased. Furthermore, the scale of the transformation will be increased by both increasing the amount of competent cells and of the DNA of each transformation.

Objective 2:
We will clone a large set of identifiable HSPs in order to generate a focused HSP library. This library will then be used to identify combinations of HSPs which might be able to generate the tolerant phenotype. The MultiSite Gateway® Pro system (Invitrogen) will be used to randomly clone three HSPs into a single expression plasmid that only requires one antibiotic for maintenance. HSP genes (listed in the proposal) from *E. coli* and *L. plantarum* based on literature and database mining will be PCR amplified and cloned into the entry clones. Then, the clones of the three entry plasmids will be combined with each other and the destination vector to produce a final expression clone containing three random HSP genes in a single expression destination vector. Finally, the destination vectors containing the three different HSP genes will be transformed into *E. coli* (K-12 MG1655) or *E. coli* (K-12 MG1655) which already carries a a groESL over-expressing plasmid. The library carrying cells will be exposed to selective or differential conditions (exposure to a stressant vs. not; here the two alcohols discussed above, plus butanol, as proposed) with the assumption that some HSP(s) represented in the library will allow for preferential growth under the applied conditions. Because the process of generating the large HSP library is slow and tedious, we will also examine combinations of select HSPs with GroESL overexpression. We will chose these GroESL "partners" based on genetic and biophysical studies, aiming to enhance the proven effectiveness of the GroESL overexpression in generating tolerance to chemicals. Such partners will include *dnaK* and small HSPs known to interact with GroESL. We will also examine the impact of expressing the *groESL* genes from a strong inducible promoter versus the natural promoter used so far. Finally, if time allows, we will consider expressing the *groESL* genes alone or with partners in ethanologenic *E. coli* cells to examine ethanol tolerance under more realistic bioprocessing conditions.

f. Major Problems/Issues (if any).

None was encountered that we could not resolve without deviating from our set goals.

g. Technology Transfer

While we have filed an Invention Disclosure and Provisional patent (see under **Productivity**), we have not yet reached the point of engaging in any Technology Transfer activities.

f. Foreign Collaborations and Supported Foreign Nationals

No foreign collaborations. Because we could not recruit experienced graduate students immediately when the funding decision for this grant was made last summer, the best approach for making progress was to hire a postdoctoral fellow (Dr. Changhao Bi; a citizen of China (foreign national); Dr. Bi works on Objective 1) who had applied for a position in our lab, and then recruit a first year PhD student (Mr. Kyle Zingaro, a US citizen who works on Objective 2). Furthermore, a senior PhD student (Mr. Sergios Nicolaou; a citizen of Cyprus (foreign national)) has been involved with this project since it conception almost 2 years ago. Mr. Nicolaou has provided methods guidance and intellectual support to both Dr. Bi and Mr. Zingaro, although he has not been financially supported by this ONR grant in year 1.